AEM
Journals.ASM.org

Differentiation of *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis* on the Basis of the *csaB* Gene Reflects Host Source

Jinshui Zheng,[a] Donghai Peng,[a] Xiaoling Song,[a] Lifang Ruan,[a] Jacques Mahillon,[b] Ming Sun[a]

State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China[a]; Laboratory of Food and Environmental Microbiology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium[b]

***csaB* gene analysis clustered 198 strains of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* into two groups related to mammalian and insect hosts, respectively. Mammal-related group I strains also have more S-layer homology (SLH) protein genes than group II strains. This indicates that *csaB*-based differentiation reflects selective pressure from animal hosts.**

Differentiation among members of the *Bacillus cereus* group has been conducted by many research groups based on single or multiple gene markers (1–10). For most of these typing analyses, *Bacillus anthracis* could be discriminated from the others, whereas *B. cereus* and *Bacillus thuringiensis* strains tended to intermingle with each other (4, 11). One reason for the difficulty in discriminating within the *B. cereus* group is that the selected marker genes are highly conserved among the tested strains. Consequently, genes or sequences with higher levels of genetic diversity should be screened and selected to better discriminate among the *B. cereus* group strains and isolates.

The surface layer (S-layer) is the outmost cell structure of many archaea and bacteria and consists of protein(s) or glycoprotein(s) (12). All S-layer proteins in *B. anthracis*, *B. cereus*, and *B. thuringiensis* display three S-layer homology (SLH) domains (PF00395). In *B. anthracis*, there are more than 20 SLH proteins, and each protein has three copies of the SLH domain (13). The product of the *csaB* gene is involved in the addition of a pyruvyl group to a peptidoglycan-associated polysaccharide fraction, a modification necessary for the binding of SLH proteins to the secondary cell wall polymer (SCWP) in some Gram-positive bacteria (14). In this study, we evaluated *csaB* as a marker for the identification and/or discrimination of *B. anthracis*, *B. cereus*, and *B. thuringiensis* strains.

When focusing on the 122 strains (21 *B. anthracis*, 79 *B. cereus* and 22 *B. thuringiensis* strains) (see Table S1 in the supplemental material) whose genome sequences are available from GenBank, we found that the *csaB* gene is located on each chromosome as a single copy. The *csaB* sequences show more diversity (nucleotide sequence identity range, 75 to 100%; average, 84%) than other marker genes, (e.g., ranges of 93% to ~100% [average, 97%] for *groEL*, 90% to ~100% [average, 97%] for *sodA*, and 86% to ~100% [average, 92%] for *gyrB*). The major topology of the phylogenetic tree based on *csaB* sequences was similar to those of *groEL*, *sodA*, and *gyrB* for the 122 selected strains (data not shown).

In addition, we amplified another 76 *csaB* fragments from *B. thuringiensis* strains (see Table S2 in the supplemental material) with primers P1 (5′-GTGCGTTTAGTCTTATCAGGAT-3′) and P2 (5′-CTTTCGCATCCCAATAMCKYACACT-3′) and sequenced the amplicons in both directions (see Text S1 in the supplemental material). An unrooted phylogenetic tree was constructed based on the above 198 sequences using the neighbor-joining algorithm implemented in MEGA5 (15) after alignment by ClustalW (16). All strains could be assembled into two groups, I and II (Fig. 1). Members of group I were subdivided into two subgroups, Ia and Ib (Fig. 1). Subgroup Ia contains 21 *B. anthracis* strains, 28 *B. cereus* strains (including 9 isolates associated with illnesses in higher animals), and 31 *B. thuringiensis* strains. Subgroup Ib contains 1 *B. thuringiensis* strain (4BQ1) and 15 *B. cereus* strains. All 6 emetic *B. cereus* strains with available *csaB* genes are located in group I; four (AH187, AND1407, H3081.97, and NC7401) in Ia and two (CER057 and CER074) in Ib. Emetic *B. cereus* produces cereulide, which causes nausea and vomiting after ingestion. Two subgroups of group II were supported by high bootstrap values. Subgroup IIa includes 4 *B. cereus* strains (MM3, R309803, BAG6X1-1, and BAG2X1-2) and *B. thuringiensis* subsp. *konkukian* strain 97-27, which was previously shown to be more closely related to *B. anthracis* (6). Higher levels of genetic conservation of the *csaB* gene were observed within subgroup IIb. Thirty out of 79 tested *B. cereus* strains and 66 out of 98 tested *B. thuringiensis* strains are distributed on 35 branches, indicating that *B. thuringiensis* strains are predominately found in subgroup IIb. Among the 96 strains in this subgroup, only 5 (172560W, AH1134, B4264, F65185, and G9842) were reported as human pathogens. *Bacillus cytotoxicus* NVH 391-98, which was isolated from an outbreak of food poisoning (17), and *B. cereus* BAG5X-1, which was isolated from soil, are not related to the two groups. Other methods also identified the former strain as an outlier (18).

The *csaB* gene plays a crucial role in the maintenance of a family of proteins covering the whole cell; therefore, its diversity could indirectly be driven by the surrounding environment. We therefore investigated ecological niches and origins of the strains in the two *csaB* clusters. More than 80% of the strains (40 out of 48 strains) isolated from humans and other mammals were clustered in group I. These strains particularly include most of the strains pathogenic to mammals: i.e., all 21 *B. anthracis* strains, 11 *B. cereus* strains associated with human or animal infection (03BB102, 95/

Received 23 February 2013 **Accepted** 28 March 2013
Published ahead of print 5 April 2013
Address correspondence to Ming Sun, m98sun@mail.hzau.edu.cn.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00591-13.

doi:10.1128/AEM.00591-13

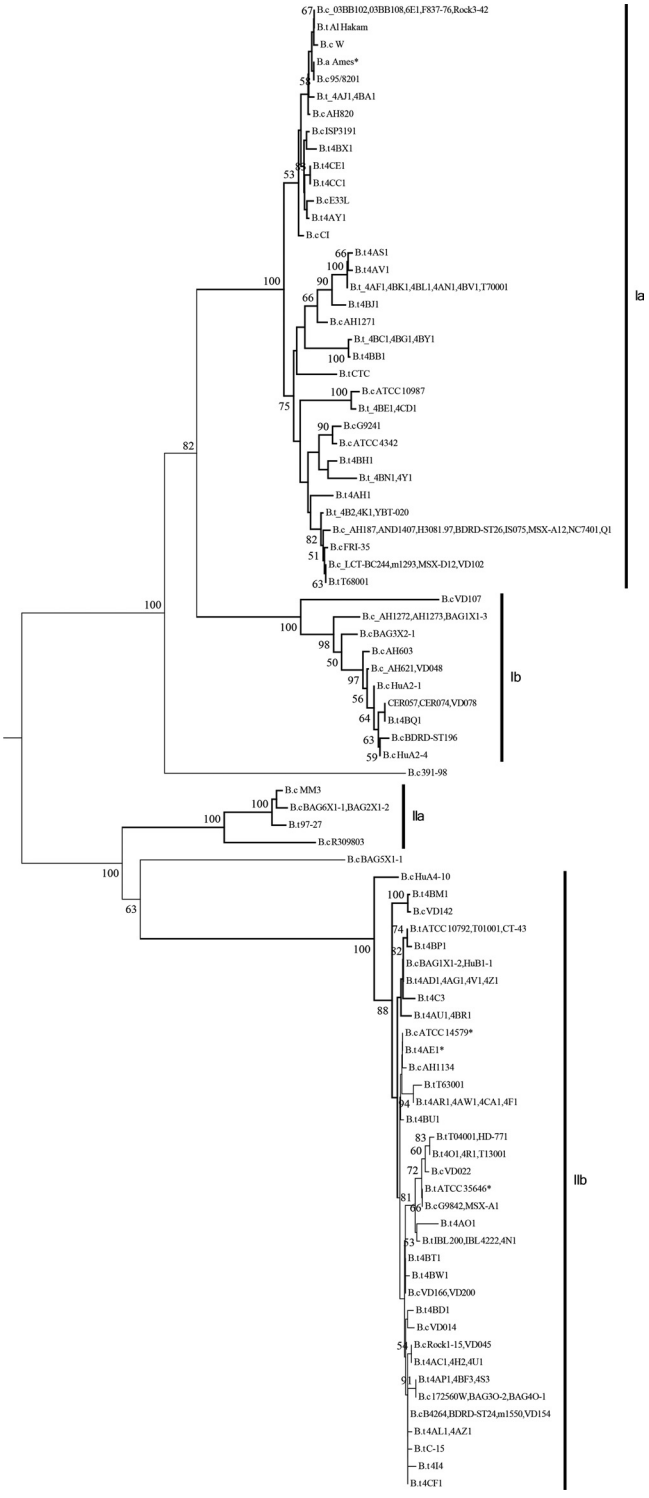

FIG 1 Unrooted phylogenetic tree of the *csaB* genes of *B. anthracis*, *B. cereus*, and *B. thuringiensis* strains. The number at each branch point represents the percentage of bootstrap support calculated from 1,000 replicates. Only bootstrap values above 50 are shown. B.a. Ames* represents all 21 *B. anthracis* strains. B.c ATCC 14579* represents *B. cereus* strains ATCC 14579, ATCC 10876, AH676, BAG3X2-2, BAG4X12-1, BDRD-Cer4, F65185, VD156, and VD169. B.t 4AE1* represents *B. thuringiensis* strains 4AE1, 4AM1, 4AQ1, 4AT1, 4CB1, 4D11, 4G1, 4L1, 4T1, 4W1, 4X1, HD-1, HD73, T03a001, and YBT-1520. B.t ATCC 35646* represents *B. thuringiensis* strains ATCC 35646, 4AK1, 4AX1, 4BS1, 4BZ1, 4M1, BMB171, Bt4Q7, HD-789, and T69001.

8201, AH1272, AH1273, AH820, E33L, F837/76, G9241, AH1271, IS075, and MSX-A12), and the 8 emetic *B. cereus* strains. Strains isolated from insects were clustered predominantly in group II. Among the 22 *B. thuringiensis* strains that were isolated from insects, only BGSC 4B2, which was isolated from *Malacosoma disstria*, is in group I. Particularly all strains of *B. thuringiensis* displaying high level of toxicity against insects and used as biopesticides fit into subgroup IIb, including *B. thuringiensis* subsp. *thuringiensis* strains T01001 and ATCC 10792, *B. thuringiensis* subsp. *kurstaki* strains 4D11, HD-1, T03a001, and YBT-1520, and *B. thuringiensis* subsp. *israelensis* strains ATCC 35646 and Bt4Q7. Accordingly, *csaB* gene sequences show a strong correlation to mammalian or insect hosts, indicating convergent evolution driven by host adaptation. However, when the 52 soilborne strains were considered, they were found to be almost equally distributed between groups I (24 strains) and II (28 strains). Soil does not appear to exert a major selective pressure for sequence diversity of *csaB* in strains of the *B. cereus* group. However, soil is a reservoir for both insect pathogens and strains pathogenic to mammals, and most of the soil isolates are not pathogenic.

CsaB is essential for anchoring the SLH proteins to the SCWP. SLH proteins are located on the outermost layer of bacteria, and some of these proteins mediate bacterial adhesion to their host (19, 20). Therefore, the distribution and evolution of all the SLH proteins present in the 122 available *B. cereus* group strains (see Table S1 in the supplemental material) were investigated. SLH protein sequences were retrieved using HMMER software with the hmmsearch command (21) and the SLH domain model PF00395 (Fig. 2; see Table S3 in the supplemental material). Strains in group I possess the most abundant SLH protein genes, with up to 26 genes in strains *B. cereus* 03BB108 and *B. thuringiensis* BGSC 4AJ1 and 4CC1, while the group II strains harbor significantly fewer ($P < 2.2e{-}16$, Mann-Whitney test). The smallest number (8 SLH protein genes) was found in several strains in subgroup IIb. For function prediction, all of the SLH proteins were searched against the Conserved Domain Database (CDD) (22) and could be classified into 13 categories (see Table S3). SLH proteins with basal metabolism function, such as those involved in peptidoglycan catabolism, are conserved and distributed in all strains among groups I and II. In contrast, SLH proteins not involved in basal metabolism show more diversity and a narrower distribution among the two groups. Most SLH proteins which contribute to the bacterial adaptation to higher animal hosts are contained by group I strains. For instance, S-layer proteins have been reported to mediate bacterial resistance against bactericidal complement activity, to participate in the adhesion to extracellular matrix proteins, and to temper the proinflammatory cytokine response (12, 23). The genes encoding these proteins are mainly contained by strains in group I but not group II (see Table S3). The gene encoding adhesion protein BlsA, which mediates adherence of the vegetative form of the *B. anthracis* strain to human cells (19), was only found in several *B. cereus* strains in group I (G9241, biovar *anthracis* strain CI, and 03BB102) which exhibit *B. anthracis*-like characteristics (24). Other examples included genes coding for the Ca^{2+} binding domain proteins, which are mainly found in group I and subgroup IIa strains, and the lactamase proteins that occur in strains of subgroup Ia. Their products are also usually involved in bacterial adaptation to their hosts.

Taken together, these observations indicate that the two *csaB*-

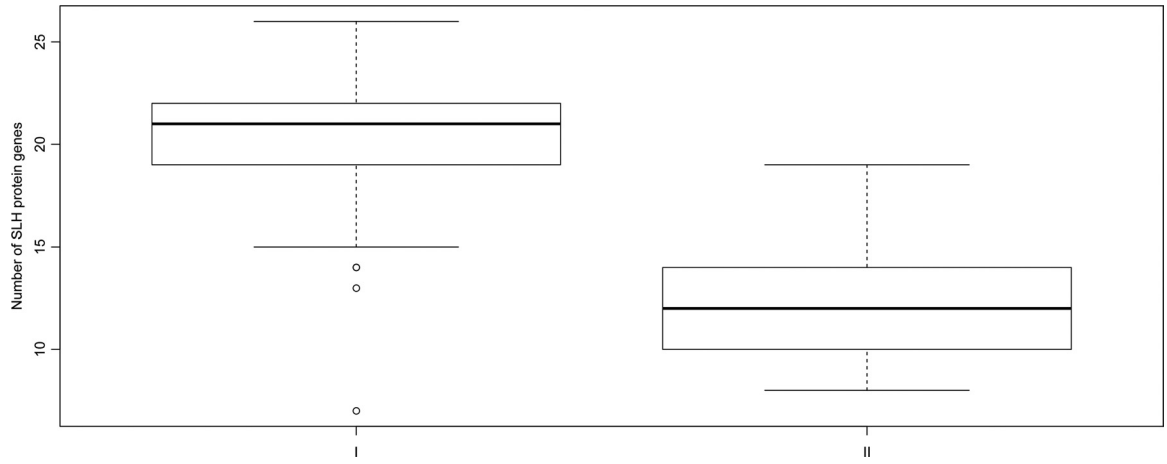

FIG 2 Summary of SLH protein gene distribution. The numbers I and II refer to the two groups shown on the *csaB* phylogenetic tree in Fig. 1. In the boxes, the bold line represents the median, and the upper and lower boundaries represent the 75th and 25th percentiles, respectively. Group I strains have significantly more SLH protein genes than group II strains ($P < 2.2e-16$, Mann-Whitney test).

based groups correspond to two distinct lifestyle environments—higher animals and insects for groups I and II, respectively.

ACKNOWLEDGMENTS

This work was supported by grants from the National High Technology Research and Development Program (863) of China (2011AA10A203), the National Basic Research Program (973) of China (2009CB118902), China 948 Program of Ministry of Agriculture (2011-G25), and the National Natural Science Foundation of China (39870036 and 31270137).

We also thank A. Gillis for critical reading of the manuscript.

REFERENCES

1. **Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB.** 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. **66**:2627–2630.
2. **Shangkuan YH, Yang JF, Lin HC, Shaio MF.** 2000. Comparison of PCR-RFLP, ribotyping and ERIC-PCR for typing *Bacillus anthracis* and *Bacillus cereus* strains. J. Appl. Microbiol. **89**:452–462.
3. **Chang YH, Shangkuan YH, Lin HC, Liu HW.** 2003. PCR assay of the *groEL* gene for detection and differentiation of *Bacillus cereus* group cells. Appl. Environ. Microbiol. **69**:4502–4510.
4. **Guinebretiere MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M, De Vos P.** 2008. Ecological diversification in the *Bacillus cereus* group. Environ. Microbiol. **10**:851–865.
5. **Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, Cherni A.** 2004. Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. J. Clin. Microbiol. **42**:3711–3730.
6. **Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA, Laker M, Pardington PE, Richardson AP, Tonks M, Beecher DJ, Kemp JD, Kolsto AB, Wong AC, Keim P, Jackson PJ.** 2004. Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl. Environ. Microbiol. **70**:1068–1080.
7. **Valjevac S, Hilaire V, Lisanti O, Ramisse F, Hernandez E, Cavallo JD, Pourcel C, Vergnaud G.** 2005. Comparison of minisatellite polymorphisms in the *Bacillus cereus* complex: a simple assay for large-scale screening and identification of strains most closely related to *Bacillus anthracis*. Appl. Environ. Microbiol. **71**:6613–6623.
8. **Zhong W, Shou Y, Yoshida TM, Marrone BL.** 2007. Differentiation of *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis* by using pulsed-field gel electrophoresis. Appl. Environ. Microbiol. **73**:3446–3449.
9. **Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182**:2928–2936.
10. **Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, DeSantis T, Brodie E, Nelson K, Pei Z.** 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl. Environ. Microbiol. **76**:3886–3897.
11. **Priest FG, Barker M, Baillie LW, Holmes EC, Maiden MC.** 2004. Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. **186**:7959–7970.
12. **Sára M, Sleytr UB.** 2000. S-layer proteins. J. Bacteriol. **182**:859–868.
13. **Kern JW, Schneewind O.** 2008. BslA, a pXO1-encoded adhesin of *Bacillus anthracis*. Mol. Microbiol. **68**:504–515.
14. **Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M, Fouet A.** 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J. **19**:4473–4484.
15. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.
16. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.
17. **Lapidus A, Goltsman E, Auger S, Galleron N, Segurens B, Dossat C, Land ML, Broussolle V, Brillard J, Guinebretiere MH, Sanchis V, Nguen-The C, Lereclus D, Richardson P, Wincker P, Weissenbach J, Ehrlich SD, Sorokin A.** 2008. Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. Chem. Biol. Interact. **171**:236–249.
18. **Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK, Sozhamannan S, Mateczun AJ, Du L, Read TD.** 2012. Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*. Genome Res. **22**:1512–1524.
19. **Kern J, Schneewind O.** 2010. BslA, the S-layer adhesin of *B. anthracis*, is a virulence factor for anthrax pathogenesis. Mol. Microbiol. **75**:324–332.
20. **Daou N, Buisson C, Gohar M, Vidic J, Bierne H, Kallassy M, Lereclus D, Nielsen-LeRoux C.** 2009. IlsA, a unique surface protein of *Bacillus cereus* required for iron acquisition from heme, hemoglobin and ferritin. PLoS Pathog. **5**:e1000675. doi:10.1371/journal.ppat.1000675.
21. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. **39**:W29–W37.
22. **Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N,**

Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. **39**:D225–D229.

23. **Mignot T, Mesnage S, Couture-Tosi E, Mock M, Fouet A.** 2002. Developmental switch of S-layer protein synthesis in *Bacillus anthracis*. Mol. Microbiol. **43**:1615–1627.

24. **Forsberg LS, Choudhury B, Leoff C, Marston CK, Hoffmaster AR, Saile E, Quinn CP, Kannenberg EL, Carlson RW.** 2011. Secondary cell wall polysaccharides from *Bacillus cereus* strains G9241, 03BB87, and 03BB102 causing fatal pneumonia share similar glycosyl structures with the polysaccharides from *Bacillus anthracis*. Glycobiology **21**: 934–948.